DTU Library

Wind Noise Reduction Using Non-negative Sparse Coding

Schmidt, Mikkel N.; Larsen, Jan; Hsiao, Fu-Tien

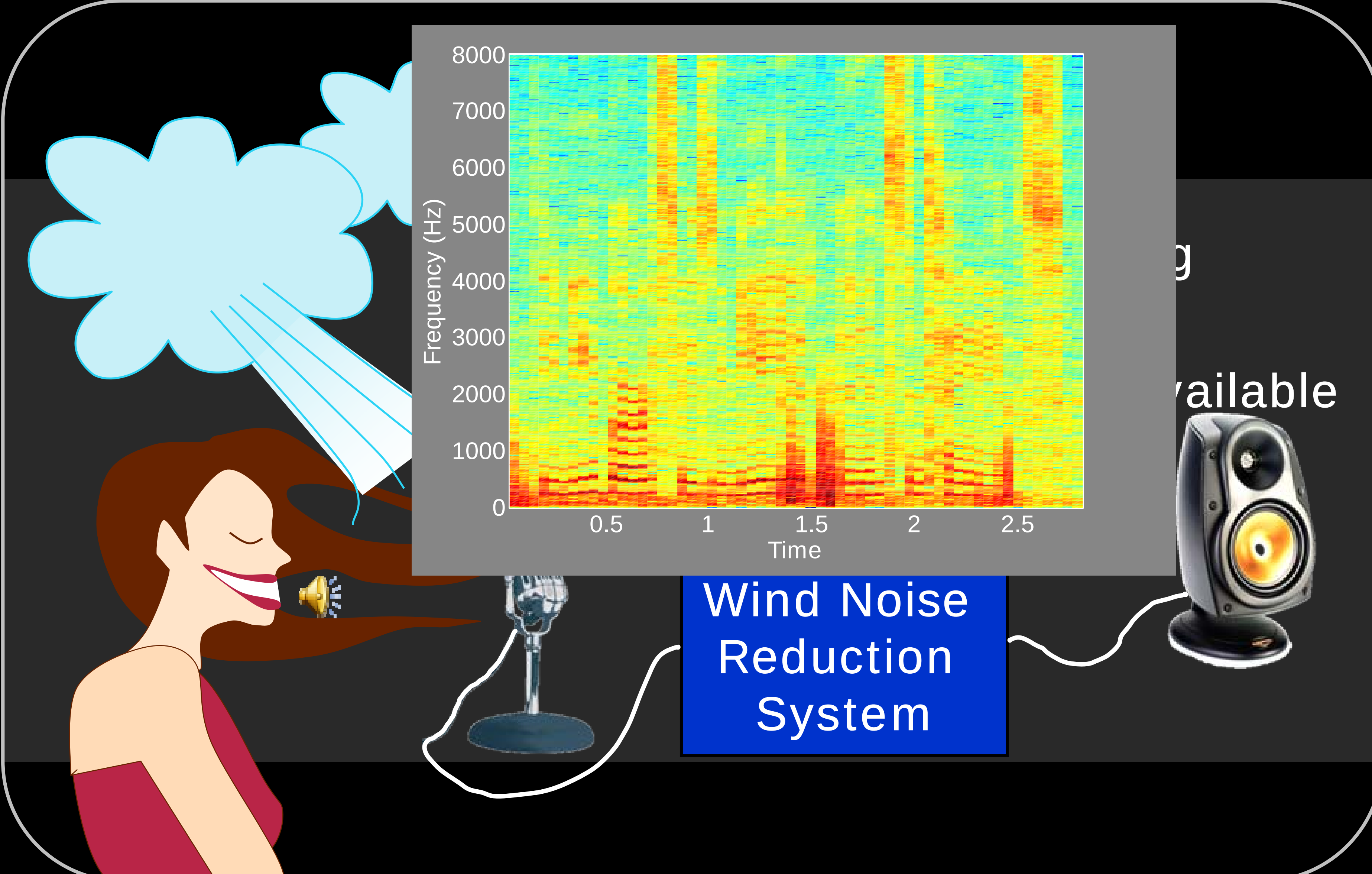
Publication date:
2007

Citation (APA):
Schmidt, M. N. (Author), Larsen, J. (Author), & Hsiao, F-T. (Author). (2007). Wind Noise Reduction Using Non-negative Sparse Coding. Sound/Visual production (digital)

www.auntiegravity.co.uk

Wind Noise Reduction Using Non-negative Sparse Coding

Mikkel N. Schmidt, Jan Larsen, Technical University of Denmark
Fu-Tien Hsiao, IT University of Copenhagen


8000
7000
6000
5000
4000
3000
2000
1000
0
Frequency (Hz)
0.5
1
1.5
2
2.5
Time
Wind Noise Reduction System

The spectrum of alternative methods

- Wiener filter (Wiener, 1949)
- Spectral subtraction (Boll 1979; Berouti et al. 1979)
- AR codebook-based spectral subtraction (Kuropatwinski & Kleijn 2001)
- Minimum statistics (Martin et al. 2001, 2005)
- Masking techniques (Wang; Weiss & Ellis 2006)
- Factorial models (Roweis 2000,2003)
- MMSE (Radfar&Dansereau, 2007)
- Non-negative sparse coding (Schmidt & Olsson 2006)

Noise Reduction

Estimate the speaker, $s(t)$, given a noisy recording $x(t)$

$$x(t) = s(t) + n(t)$$

... based on prior knowledge of the noise, $n(t)$

Single Channel Source Separation

Hard problem: There is no spatial information

- we cannot use
  - Beamforming
  - Independent component analysis

Signal Representation

- Exponentiated magnitude spectrogram

$$\boldsymbol{X} = |\mathrm{STFT}\{x(t)\}|^{\gamma}$$

$\gamma = 2$	Power spectrogram
$\gamma = 1$	Magnitude spectrogram
$\gamma = 0.67$	Cube root compression
	(Steven's power law - perceived intensity)

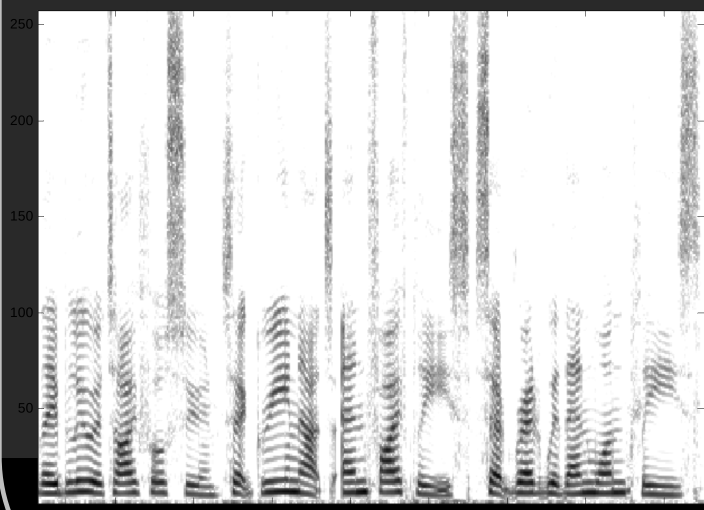
- Ignore phase information. Reconstruct by re-filtering

Non-negative Sparse Coding

- Factorize the signal matrix

$$X \approx DH$$

Spectrogram

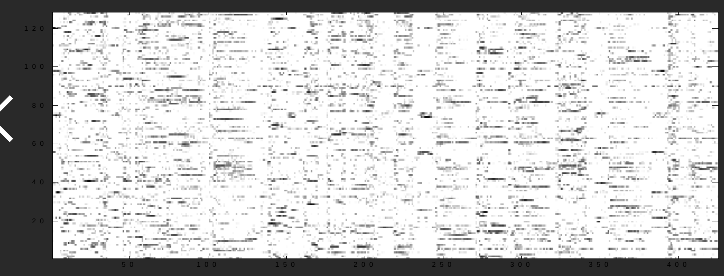


$\approx$

Dictionary

$\times$

Sparse Code



DTU

Non-negative Sparse Coding

- Factorize the signal matrix

$$X \approx DH$$

where D and H are non-negative and H is sparse

- Non-negativity: Parts-based representation, only additive and not subtractive combinations
- Sparseness: Only few dictionary elements active simultaneously. Source specific and more unique.

The Dictionary and the Sparse Code

$$X \approx DH$$

- Dictionary, D
  - Source dependent over-complete basis
  - Learned from data

- Sparse Code, H
  - Time & amplitude for each dictionary element
  - Sparseness: Only a few dictionary elements active simultaneously

Non-negative Sparse Coding of Noisy Speech

- Assume sources are additive

$$\boldsymbol{X} = \boldsymbol{X}_s + \boldsymbol{X}_n \approx [\boldsymbol{D}_s \ \boldsymbol{D}_n] \begin{bmatrix} \boldsymbol{H}_s \\ \boldsymbol{H}_n \end{bmatrix} = \boldsymbol{D}\boldsymbol{H}$$

Permutation Ambiguity

$$X \approx [D_s \ D_n] \begin{bmatrix} H_s \\ H_n \end{bmatrix} = DH = (DP)(P^\top H)$$

- Precompute both dictionaries (Schmidt & Olsson 2006)
- Devise a grouping rule (Wang & Plumbley 2005)
- Precompute wind dictionary and learn speech dictionary from noisy recording
- Use multiplicative update rule (Eggert&Körner 2004)

Other rules could be used e.g. projected gradient (Lin, 2007)

Importance and sensitivity of parameters

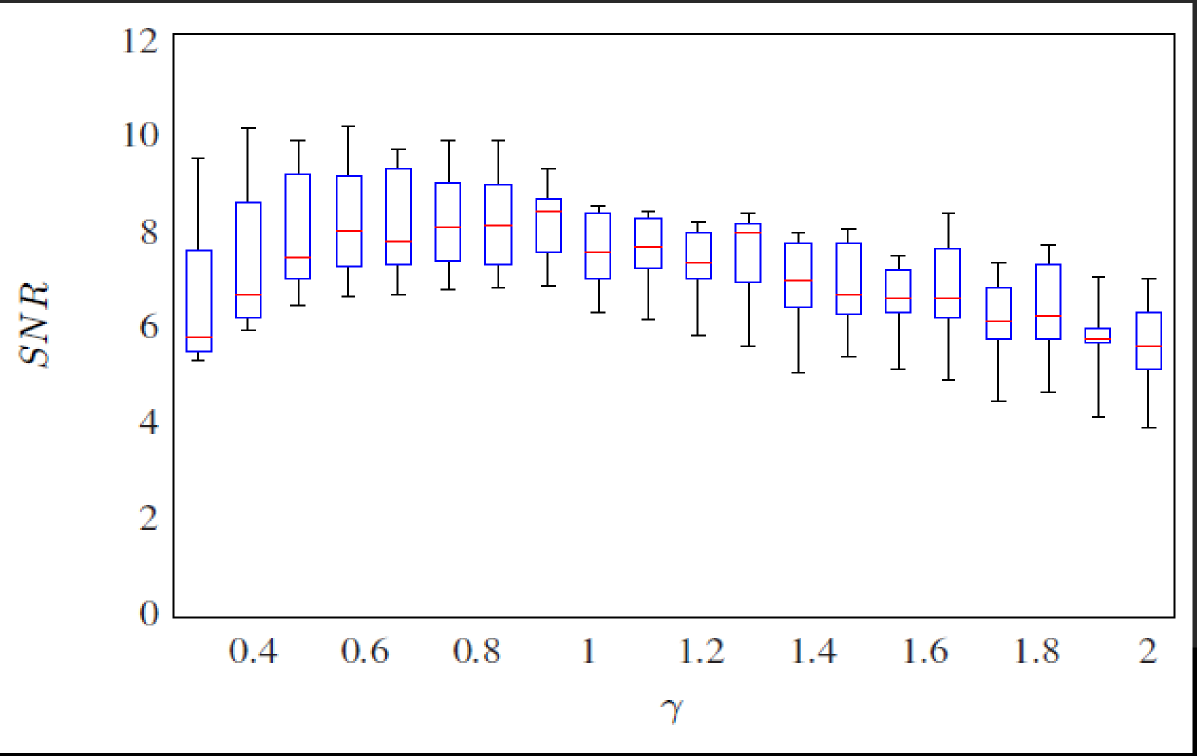
- Representation
  - STFT exponent
- Sparseness
  - Precomputed wind noise dictionary
  - Wind noise
  - Speech
- Number of dictionary elements
  - Wind noise
  - Speech

Quality Measure

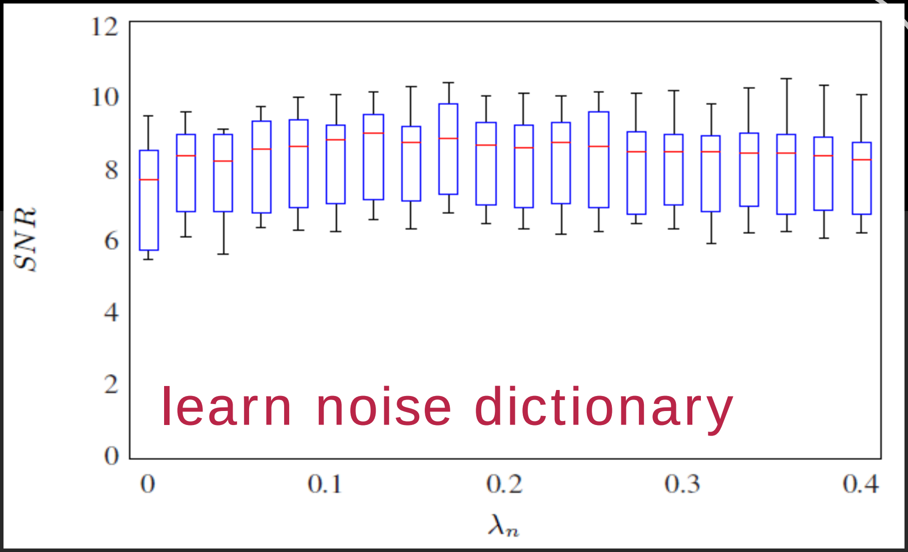
- Signal to noise ratio
  - Simple measure, has only indirect relation to perceived quality
- Representation-based metrics
  - In systems based on time-frequency masking, evaluate the masks
- Perceptual models
  - Promising to use PEAQ or PESQ
- High-level Attributes
  - For example word error rate in a speech recognition setup
- Listening-tests
  - Expensive, time-consuming, aspects (comfort, intelligibility)

Signal Representation

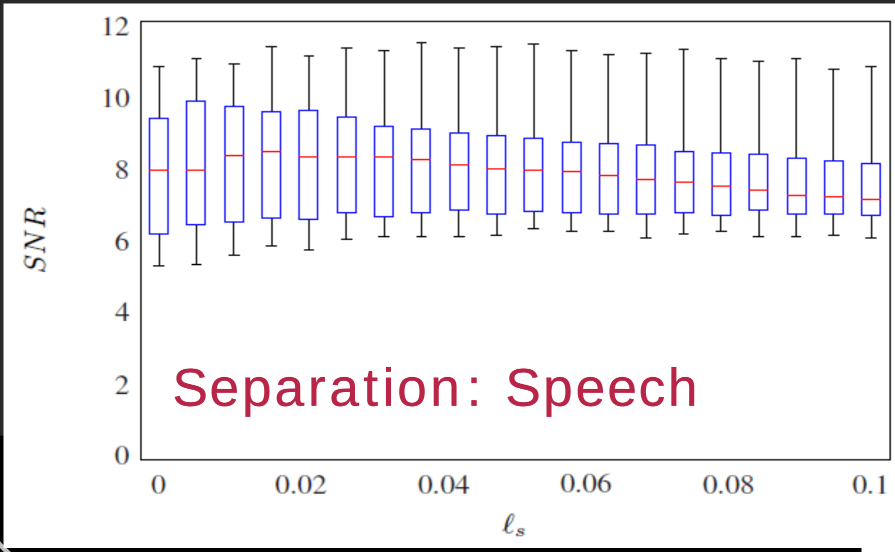
- Exponentiated magnitude spectrogram

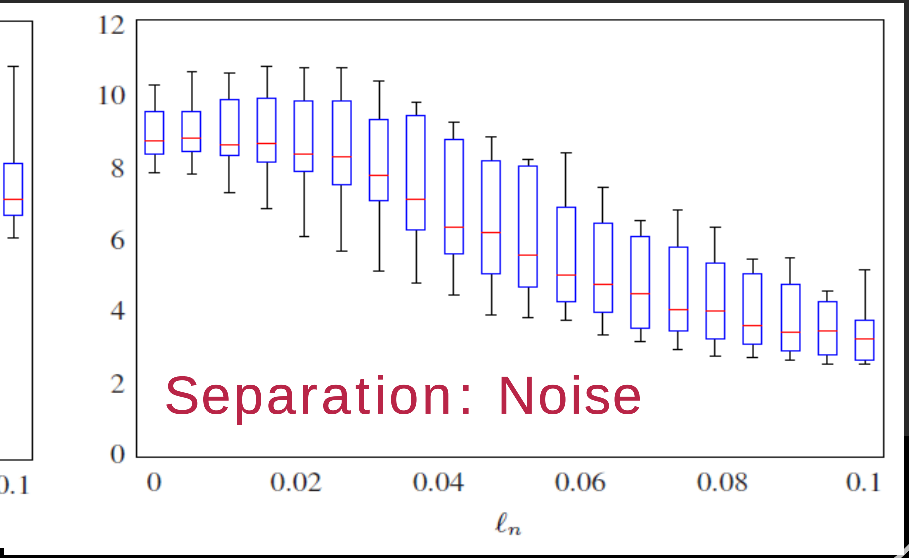


Sparseness

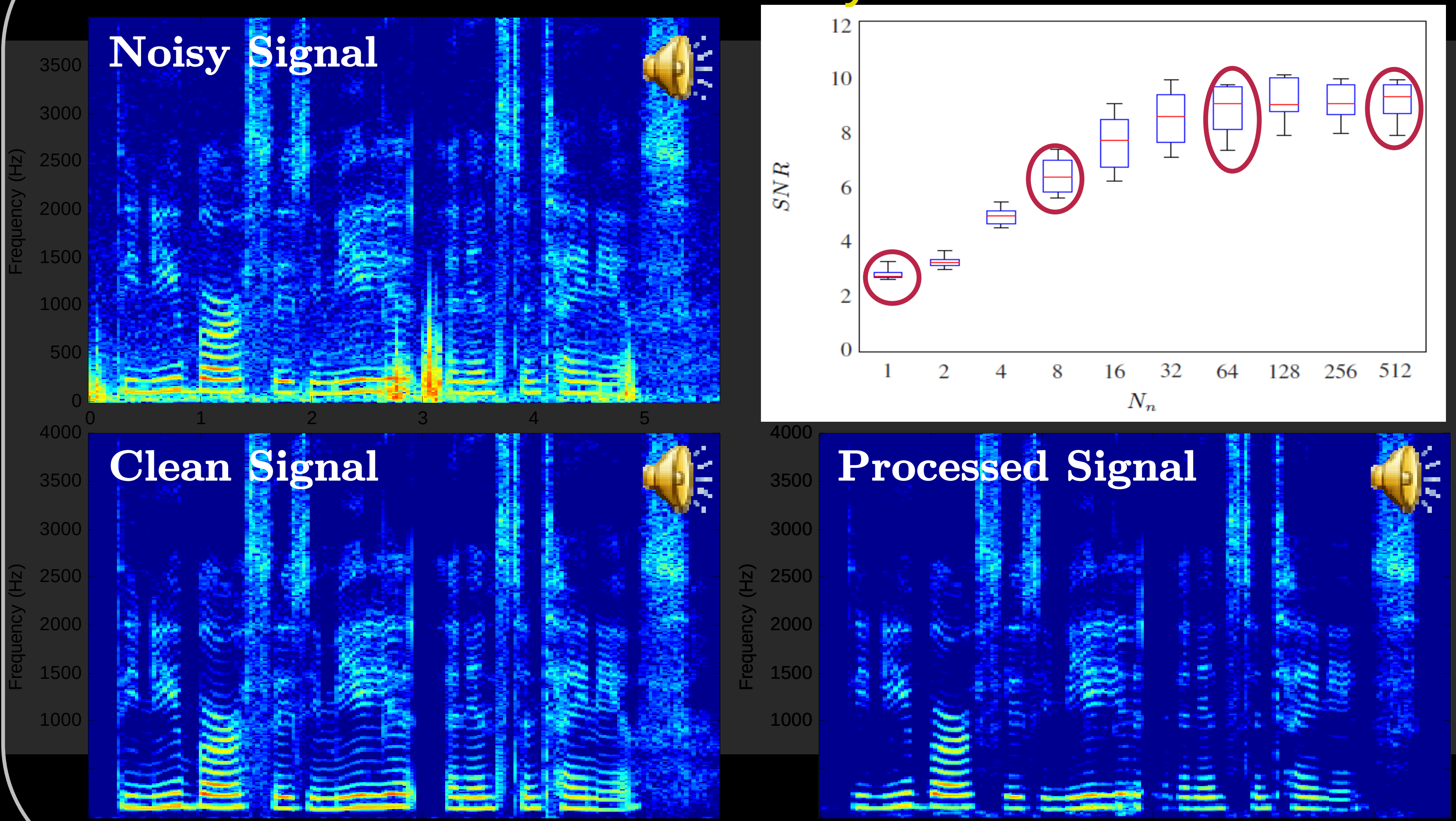


Qualitatively: Tradeoff between residual noise and speech distortion



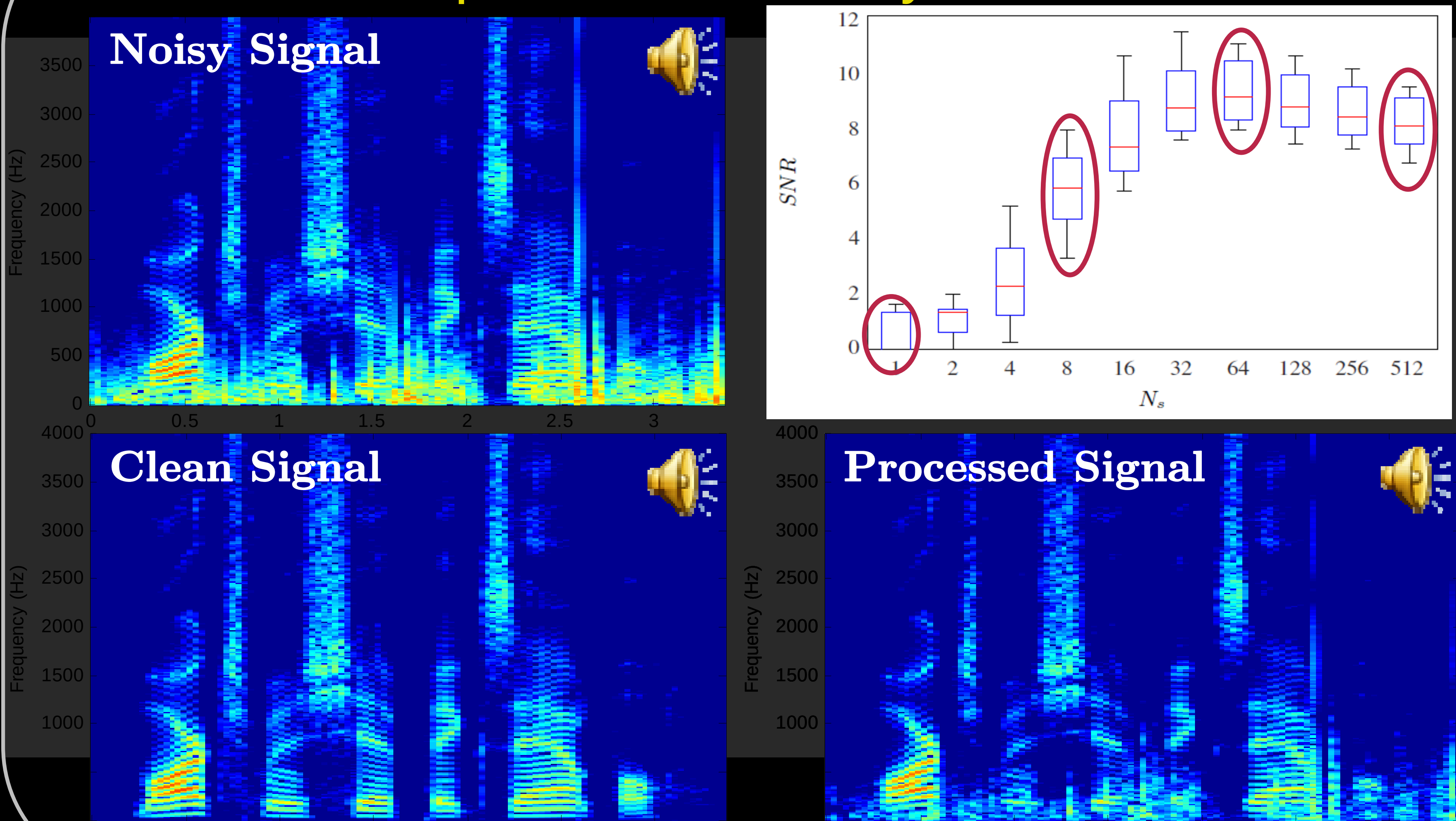


Number of Noise-Dictionary Elements



Mikkel Schmidt, Jan Larsen

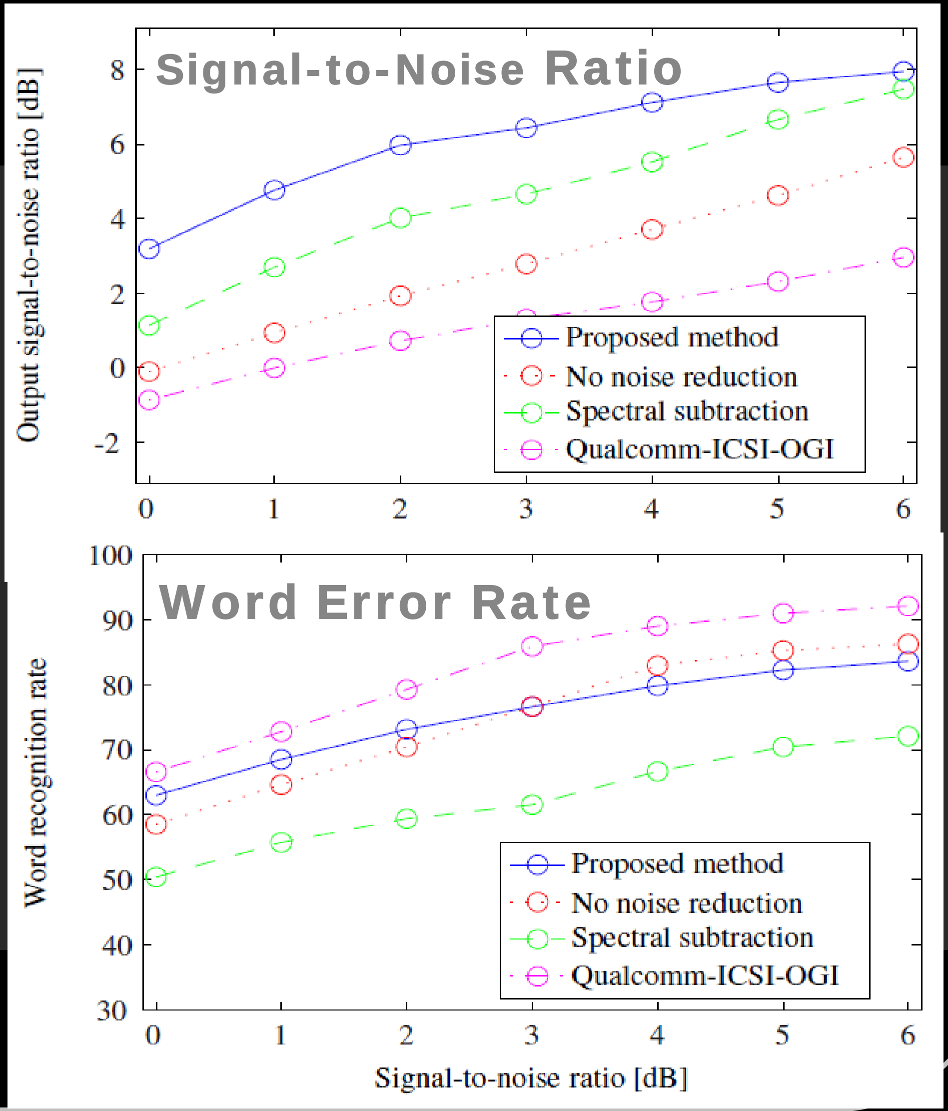
Number of Speech-Dictionary Elements



Mikkel Schmidt, Jan Larsen

Comparison

1. Proposed method
2. No noise reduction
3. Spectral subtraction
4. Qualcomm-ICSI-OGI aka adaptive Wiener filtering (Adami et al. 2002)



Conclusions and outlook

- Sparse coding of spectrogram representations is a useful tool for reduction of wind noise
- Only samples of wind noise are required

›Careful evaluation and integration of perceptual measures

›Handling nonlinear saturation effects

›Optimization of performance (fewer freq. bands, adaptation to new situations)